DOCUMENT RESUME

ED 052 138 SP 004 983

AUTHOR	Rosenshine, Barak; And Others
TITLE	Empirical Validation of a New Category System: One Example.
INSTITUTION	Temple Univ., Philadelphia, Pa.
PUB DATE	71
NOTE	9p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Academic Achievement, *Classroom Observation Techniques, Elementary School Students, Grade 3, *Interaction Process Analysis, *Predictive Validity, *Teacher Behavior
IDENTIFIERS	Expanded Interaction Analysis System, Flanders System of Interaction Analysis

ABSTRACT

This study found that data from previous research can be used to validate a new observational category system and that subscripting of the original ten categories of the Flanders Interaction Analysis System is useful in identifying more specific behaviors which correlate with student achievement. The new category system was the Expanded Interaction Analysis System (EIAS), in which each of the 10 Flanders categories is subscripted into two to four smaller categories. The material used for validation was the audiotapes, typescripts, and residual student achievement scores used in the original study by Wright and Nuthall (1970). In this study, 17 teachers presented three 10-minute lessons on the black-billed gull to intact classes of third grade students. Results from the EIAS coding and analysis showed the major advantage of subscripting to be in categories 3 and 4. In category 3 (teacher accepts students' ideas), the entire category had a correlation of .15 with achievement, whereas the subscript 3a (acknowledges ideas by simple reflection) yielded a correlation of .30. In category 4, although the total frequency of questions yielded a correlation of .18, the asking of factual questions had a correlation of .55 with student achievement. This was considered to be due to the factual nature of the material. This technique of using the same data with different systems is recommended for future use. (RT)

Barak Rosenshine
Edward Crill
Richard McAdams Temple Univ.

EMPIRICAL VALIDATION OF A NEW CATEGORY SYSTEM: ONE EXAMPLE

Objective

New observational category systems are continually being developed, and these new systems are supported by arguments for their face validity, their value in describing teaching, or their ability to discriminate between teachers rated as "outstanding" or "poor" by principals, supervisors, and/or students. However, few tests have been made of whether the variables in a new category system are useful in predicting student achievement. This study is an example of one such test.

Procedures

The new category system chosen for study was the Expanded Interaction Analysis System (EIAS) developed by E. Amidon and P. Amidon. It was chosen because one of the investigators had been trained in its use.

The EIAS is one of at least 20 category systems which have been developed from the original, 10-category system by Flanders. All these systems reflect the concern that the 10 original categories seem too gross, and that attention should be given to specific behaviors within some categories. For example, Category 6 (Gives directions) has been subdivided into "Gives cognitive directions" and "Gives managerial directions." Flanders, who has used such procedures in his recent research, has labeled

the procedures "subscripting" because the original category number is retained, but subscripts are added to each number to identify the specific behaviors within each category.

In the system developed by the Amidons, each of the 10 categories in the Flanders system is subscripted into two to four smaller categories. Any set of subscripts, combined, contains all the behaviors coded within the larger category; thus, the categories as wholes and the subscripts can be used separately in an analysis.

The material for validation was the audiotapes, typescripts, and residual student achievement scores used in the original study by Wright and Nuthall (1970). (NB: The Wright and Nuthall article will be published in the November, 1970, issue of AERJ, and, therefore, will be available to members of the audience and readers.) In this study, 17 teachers presented three 10-minute lessons on the black-billed gull to intact classes of third grade students. All teachers taught from identical original material, and the residual gain scores were obtained from using two measures of aptitude as predictor variables. The audiotapes and typescripts were coded by a system developed by Wright and Nuthall, but were made available to the investigators for subsequent use.

Two of the investigators coded all of the lessons using the EIAS. Interrater reliability of .80 or better was established before coding the lessons, and neither of the coders has ever

known the residual gain scores attained by any teacher.

The results of the analysis are presented in Table 1, which contains the names of the subscripts, the frequencies of teacher behavior in each of the major categories and in the subcategories, and the correlations between each variable and the mean class residual gain scores. Both the category frequency and the subscript frequencies are presented, so that within each category the total frequencies of the subscripts should equal the total frequency in the category (allowing for rounding).

When reviewing the table and interpreting the results, one should consider the specific type of lesson. The lessons were highly factual, and the criterion test focused exclusively upon the learning of new facts and concepts about the black-billed gull.

Results

The value of subscripting. In this sample, the major advantage of subscripting using the Amidons' system lies in Categories 3 and 4. In Category 3 (Teacher accepts students' ideas), the entire category had a correlation of .15 with achievement, whereas the subscript 3a (Acknowledges ideas by simple reflection) yielded a correlation of .30[1]. The superiority of variable 3a, even in this sample, is unexpected. Investigators who discuss

Category 3 usually point to clarification of student ideas as the purer form of the variable; yet, such behavior (coded as 3b) yielded a correlation of -.08 with the criteria.

The value of subscripting appears strongest in Category 4 (Asks questions). Although the total frequency of questions yielded a correlation of .18, the asking of factual questions (4f) had a correlation of .55 with student achievement. Other types of questions yielded smaller, and sometimes negative, correlations. The high correlation for 4f in this sample appears to reflect the factual nature of the material; the other results suggest that for the type of learning required on the criterion test, the asking of divergent and evaluative questions had negative effects. These results corroborate the findings of Wright and Nuthall.

Employing subdivisions for Category 9 (Unpredictable student talk) was not as useful as leaving the category intact. In category 9 the magnitude of the correlations for each of the subdivisions was smaller than that obtained for the total category. These results illustrate the possibility that for some variables, a breakdown into smaller categories reduces the predictability of each of the smaller components.

There are also examples in Categories 2 and 6 of subscripts yielding higher correlations than the entire category, but the frequencies of the subscripted variables with the higher correlations (2p and 6m) are too small to indicate anything about the importance of these variables in classroom instruction. For the

remaining categories (Categories 1, 5, 7, 8, and 10), there does not appear to be any advantage to subscripting in this study.

Other units of analysis. The observational category system developed by Flanders, and those which derive from it, all use time (a three-second interval) as the primary unit of analysis. Other investigators have used "cognitive units" whose length depends upon the nature of the discourse taking place. Wright and Nuthall have used an "utterance," which they defined as "a complete statement by a teacher or student at any one time in the discourse." By this definition, it is possible for a teacher's utterance to contain more than one question. For example, a teacher might ask a question, pause, receive no answer, and proceed to ask a second and third question. Using this concept in their analysis, Wright and Nuthall obtained the following results:

utternaces with one question	$\underline{r}$ = .54*
utternaces with two questions	$\underline{r}$ = -.42
utterances with more than two questions	$\underline{r}$ = -.43*

Wright and Nuthall interpreted these results as indicating that a teacher who is able to develop questions which can be answered the first time he asks them is more successful (on this type of task) than a teacher who needs to repeat or rephrase his question until he receives a response. These findings and the plausible interpretation suggest that some valuable information may be obscured if time alone is used as the unit of analysis.

Significance of the Differences. Statistical analyses were also made to determine if the correlations for subdivisions within a category were significantly different from each other, or if the correlations for a category total and a category subdivision were significantly different.

When the study was conceived, we expected that differences of 25 or 30 points would be statistically significant. In fact, because of the size of the sample (N = 17) differences between correlations had to be 55 points or more before significant results were obtained. Thus, none of the correlations in Category 3 (Use of student ideas) are significantly different. Significant differences were obtained only in Categories 4 and 9.

The need for absolute differences of 55 points of more in order to obtain significant differences is one limitation of this procedure. An investigator would need a sample of 30 teachers or more in order to obtain significant differences between two correlations whose absolute difference was 30 points or so.

Discussion

In place of the deluge of new category systems which are used only to describe teaching, we propose that the validity of some of the concepts in these new systems be tested by determining how well they predict student achievement. The above example shows how such tests might be conducted. It is also very useful if different investigators conduct their tests on the same data (in this case, transcripts, tape recordings, and residual gain scores). When several category systems are tested with the same data, we can obtain more information on the predictive validity of the concepts being used in each system.

Table 1

Correlations Between Variables in the Expanded Interaction Analysis System and Residual Student Achievement

Category Number and Description	Mean Frequency	Frequency r	Percentage r
1 Accepts student feelings (Total)	2.35	.08	.18
1a- acknowledges feelings	2.23	.10	19
1c- clarifies feelings	0.12	-.16	-.15
1r- refers to similar feelings of others	0.00	--	--
2 Praises (Total)	45.06	.04	.11
2w- Praises with no criteria	43.88	.05	.13
2P- Praises with public criteria	1.18	-.31	-.30
2p- Praises with private criteria	00.00	--	--
3 Accepts student ideas (Total)	87.09	.15	.11
3a- Acknowledge ideas by simple reflection	36.00	.30	.32
3c- Clarifies ideas	36.88	- 08	-.06
3s- Summarizes ideas	14.18	.03	.02
4 Asks Questions (Total)	151.71	.18	.16
4f- Asks factual questions	83.94	.55*	.50
4c- Asks convergent questions	22.82	.14	.12
4d- Asks divergent questions	37.53	-.36	-.36
4e- Asks evaluative questions	7.06	-.34	-.27
5 Lectures (Total)	102.65	.03	.01
5f- Factual lecture	84.18	.03	.00
5m- Motivational lecture	3.00	-.17	-.06
5o- Orientation lecture	11.41	.05	.11
5p- Personal opinion lecture	4.06	.05	-.04
6 Gives directions (Total)	6.41	.07	-.12
6c- Gives cognitive directions	5.24	- 09	-.23
6m- Gives managerial directions	1.18	.41[a]	.27
7 Criticism (Total)	10.59	.18	.20
7w- Criticizes with no criteria	5.88	-.01	-.02
7P- Criticizes with public criteria	4.71	.19	.07
7p- Criticizes with private criteria	0.06	-.07	-.07

Table 1, cont'd.

Category Number and Description	Frequency	Frequency r	Percentage r
8 Predictable student talk (Total)	115.47	.45[a]	.38
8f- Factual student talk	91.65	.40	.34
8c- Convergent student talk	23.82	.23	.22
9 Unpredictable student talk (Total)	60.65	-.47[a]	-.48
9d- Divergent student response	42.24	-.36	-.38
9e- Evaluative student response	7.82	-.35	-.31
9i- Student initated talk	10.59	.18	.25
10 Silence or Confusion (Total)	21.71	.12	.11
10s- Silence	6.61	.01	-.04
10c- Confusion	15.09	.12	.14
i/d ratio	9.75	.10	
Total Tallies	611.53	.13	

*** $\underline{r}$ (.01) = .61
** $\underline{r}$ (.02) = .56
* $\underline{r}$ (.05) = .48
[a] $\underline{r}$ (.10) = .41